



# CANCER DISCOVERY

## Comprehensive Genomic Analysis of Rhabdomyosarcoma Reveals a Landscape of Alterations Affecting a Common Genetic Axis in Fusion-Positive and Fusion-Negative Tumors

Jack F. Shern, Li Chen, Juliann Chmielecki, et al.

*Cancer Discovery* Published OnlineFirst January 16, 2014.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/2159-8290.CD-13-0639](https://doi.org/10.1158/2159-8290.CD-13-0639)

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, contact the AACR Publications Department at [permissions@aacr.org](mailto:permissions@aacr.org).

## RESEARCH ARTICLE

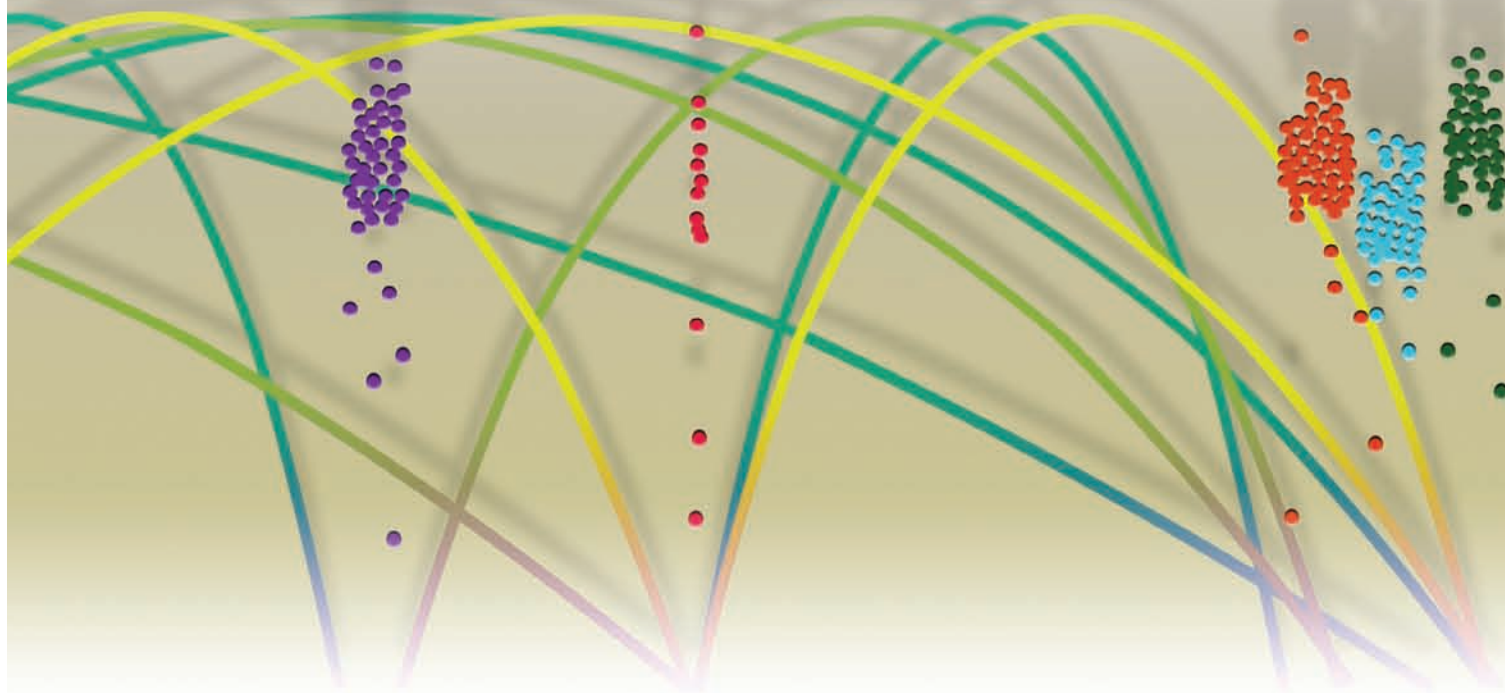
# Comprehensive Genomic Analysis of Rhabdomyosarcoma Reveals a Landscape of Alterations Affecting a Common Genetic Axis in Fusion-Positive and Fusion-Negative Tumors

Jack F. Shern<sup>1</sup>, Li Chen<sup>1</sup>, Juliann Chmielecki<sup>3,4</sup>, Jun S. Wei<sup>1</sup>, Rajesh Patidar<sup>1</sup>, Mara Rosenberg<sup>3</sup>, Lauren Ambrogio<sup>3</sup>, Daniel Auclair<sup>3</sup>, Jianjun Wang<sup>1</sup>, Young K. Song<sup>1</sup>, Catherine Tolman<sup>1</sup>, Laura Hurd<sup>1</sup>, Hongling Liao<sup>1</sup>, Shile Zhang<sup>1</sup>, Dominik Bogen<sup>1</sup>, Andrew S. Brohl<sup>1</sup>, Sivasish Sindiri<sup>1</sup>, Daniel Catchpoole<sup>9</sup>, Thomas Badgett<sup>1</sup>, Gad Getz<sup>3</sup>, Jaume Mora<sup>10</sup>, James R. Anderson<sup>6</sup>, Stephen X. Skapek<sup>7</sup>, Frederic G. Barr<sup>2</sup>, Matthew Meyerson<sup>3,4,5</sup>, Douglas S. Hawkins<sup>8</sup>, and Javed Khan<sup>1</sup>

**ABSTRACT**

Despite gains in survival, outcomes for patients with metastatic or recurrent rhabdomyosarcoma remain dismal. In a collaboration between the National Cancer Institute, Children's Oncology Group, and Broad Institute, we performed whole-genome, whole-exome, and transcriptome sequencing to characterize the landscape of somatic alterations in 147 tumor/normal pairs. Two genotypes are evident in rhabdomyosarcoma tumors: those characterized by the *PAX3* or *PAX7* fusion and those that lack these fusions but harbor mutations in key signaling pathways. The overall burden of somatic mutations in rhabdomyosarcoma is relatively low, especially in tumors that harbor a *PAX3/7* gene fusion. In addition to previously reported mutations in *NRAS*, *KRAS*, *HRAS*, *FGFR4*, *PIK3CA*, and *CTNNB1*, we found novel recurrent mutations in *FBXW7* and *BCOR*, providing potential new avenues for therapeutic intervention. Furthermore, alteration of the receptor tyrosine kinase/*RAS*/*PIK3CA* axis affects 93% of cases, providing a framework for genomics-directed therapies that might improve outcomes for patients with rhabdomyosarcoma.

**SIGNIFICANCE:** This is the most comprehensive genomic analysis of rhabdomyosarcoma to date. Despite a relatively low mutation rate, multiple genes were recurrently altered, including *NRAS*, *KRAS*, *HRAS*, *FGFR4*, *PIK3CA*, *CTNNB1*, *FBXW7*, and *BCOR*. In addition, a majority of rhabdomyosarcoma tumors alter the receptor tyrosine kinase/*RAS*/*PIK3CA* axis, providing an opportunity for genomics-guided intervention. *Cancer Discov*; 4(2); 1–16. ©2014 AACR.



## INTRODUCTION

Rhabdomyosarcoma is a myogenic cancer that is the most common soft-tissue sarcoma of childhood (1). With the development of multimodal chemotherapy regimens, relapse-free survival rates have improved to 70% to 80% in patients with localized disease, albeit with significant toxicity (2). Unfortunately, despite aggressive therapy, the 5-year survival rate for patients with metastatic disease remains only 30% (3). Currently, rhabdomyosarcoma tumors are classified by histology into two major subtypes, alveolar rhabdomyosarcoma and embryonal rhabdomyosarcoma, which have distinct

molecular and clinical profiles. Alveolar rhabdomyosarcoma carries a poor prognosis and tends to occur in adolescents. Genetically, alveolar rhabdomyosarcoma is defined in the majority of cases by a characteristic fusion between the *PAX3* or *PAX7* and *FOXO1* genes (reviewed in ref. 4). The embryonal rhabdomyosarcoma subtype typically affects younger children and portends a good prognosis when localized. Previous reports have identified a wide range of genetic aberrations in embryonal rhabdomyosarcoma, including LOH at 11p15.5 (5) as well as mutations in *TP53* (6), *NRAS*, *KRAS*, *HRAS* (7), *PIK3CA*, *CTNNB1* (8), and *FGFR4* (9).

**Authors' Affiliations:** <sup>1</sup>Pediatric Oncology Branch, Oncogenomics Section, Center for Cancer Research, NIH; <sup>2</sup>Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland; <sup>3</sup>Broad Institute of MIT and Harvard, Cambridge; <sup>4</sup>Medical Oncology and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute; <sup>5</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts; <sup>6</sup>University of Nebraska Medical Center, Omaha, Nebraska; <sup>7</sup>Department of Pediatrics, Division of Hematology/Oncology, University of Texas Southwestern Medical Center, Dallas, Texas; and <sup>8</sup>Department of Pediatrics, Seattle Children's Hospital, Fred Hutchinson Cancer Research Center, University of Washington, Seattle, Washington; <sup>9</sup>The Tumour Bank, The Children's Cancer Research Unit, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; <sup>10</sup>Department of Oncology, Hospital Sant Joan de Deu de Barcelona, Barcelona, Spain

**Note:** Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

J.F. Shern and L. Chen contributed equally to this work.

**Corresponding Author:** Javed Khan, Oncogenomics Section, Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, 37 Convent Drive, Room 2016B, Bethesda, MD 20892. Phone: 301-435-2937; Fax: 301-480-0314; E-mail: [khanjav@mail.nih.gov](mailto:khanjav@mail.nih.gov)

**doi:** 10.1158/2159-8290.CD-13-0639

©2014 American Association for Cancer Research.

## RESEARCH ARTICLE

Shern et al.

Despite an increasing understanding of the molecular mechanisms underlying these tumors, few novel agents have made their way past early-phase clinical trials, and gains in survival have mainly been made through optimization of a cytotoxic chemotherapy regimen (10). Further characterization of the genetic events underlying this tumor type is critical to the development of more effective diagnostic, prognostic, and therapeutic strategies. Here, we report a collaborative effort between the National Cancer Institute (Bethesda, MD), the Children's Oncology Group (Philadelphia, PA), and the Broad Institute (Cambridge, MA) using a combination of whole-genome sequencing (WGS), whole-exome sequencing (WES), and whole-transcriptome sequencing (WTS) along with high-resolution single-nucleotide polymorphism (SNP) arrays to characterize the landscape of somatic alterations in 147 tumor/normal pairs. Our findings describe the landscape of genetic events that occur in rhabdomyosarcoma and provide a map for future studies of targeted molecular therapies for this tumor type.

## RESULTS

A set of 44 rhabdomyosarcoma tumors with matched normal leukocyte DNA was sequenced with paired-end WGS and served as a discovery set. WGS generated an average of 294 gigabases (Gb) of sequence per sample to a mean depth of 105 $\times$ . This depth of coverage allowed high-quality calls covering 97% of the genome (Supplementary Table S1). To extend and validate our findings, we also performed WES and high-resolution SNP arrays on 103 additional tumors and their matched germlines (147 tumors in total, with clinical data summarized in Supplementary Table S2). Eighty of the tumors were analyzed by WTS, allowing us to evaluate the expression changes associated with the observed genomic alterations.

**PAX Gene Rearrangements in Rhabdomyosarcoma Tumors**

As expected, the defining genomic alteration seen across the entire cohort was recurrent t(2;13) or t(1;13) that resulted in a fusion of the N-terminus of PAX3 or PAX7 to the C-terminus of FOXO1 (refs. 11, 12; 35 had PAX3-FOXO1 and 15 had PAX7-FOXO1; Fig. 1A and B). The fusions discovered in WGS or WTS were confirmed by reverse transcription PCR (RT-PCR) when adequate RNA was available. In addition to these classic fusions, three tumors that were histologically classified as alveolar rhabdomyosarcoma, but did not have the classical PAX3/7-FOXO1 fusion by RT-PCR, were found to have alternative PAX fusions as detected by WGS or transcriptome sequencing. Cases RMS235 and RMS2031 harbored a PAX3-NCOA1 fusion that resulted from an intrachromosomal rearrangement previously described as having similar oncogenic properties as the PAX3-FOXO1 (13). We also uncovered a novel PAX fusion in a region of massive rearrangement of chromosome 2q in RMS2046 (Fig. 1C and 2A). This rearrangement resulted in a fusion of the N-terminus of PAX3 (first seven exons) and the C-terminus of INO80D, a subunit of the ATP-dependent chromatin remodeling complex. RNA sequencing of RMS2046 showed in-frame expression of the novel fusion transcript (Fig. 2B). Unsupervised clustering using the WTS data showed clear separation between tumors that harbored the rearrangement of a PAX gene from those that did not.

Of note, the tumors with the alternative PAX gene fusions clustered closely to the other alveolar rhabdomyosarcoma that harbored the classic PAX3/7-FOXO1 fusion expression profiles (Fig. 2C). Besides the three tumors that carried a novel PAX gene rearrangement, within this group there were seven additional fusion-negative alveolar histology tumors that had no PAX gene alteration but had a somatic mutation and expression profile that was more consistent with embryonal tumors (Fig. 2D and Supplementary Fig. S1A-S1D).

**Recurrent Chromosomal Structural Rearrangement in Rhabdomyosarcoma**

Beyond the rearrangement of the PAX3/PAX7 genes, WGS identified 553 somatic structural variations affecting 419 genes in 44 rhabdomyosarcoma tumor genomes (Supplementary Table S3). High-resolution SNP arrays corroborated 90% of high-confidence structural variations when a copy-number change was present (see Methods). Forty-eight genes were recurrently affected by structural variations, including genes previously implicated in rhabdomyosarcoma pathology (MIR17HG, CNR1, and CDKN2A; refs. 14-16), tyrosine kinase signaling (ERBB4, RPTOR, FRS2, and CACNA1A), and muscle development (NRG1 and FOXP2; Supplementary Table S4). Frequently (341 of 553; 61%), junction events occurred in areas of complex rearrangement or tandem duplications most often associated with regions of high copy-number amplification. Ten percent of the junctions were predicted to result in deletions. When a junction event occurred at the DNA level between two genes, a fusion transcript was produced at the RNA level 19% (55 of 296) of the time. Among these events, fusion of the PAX genes accounts for one third of the fusion transcripts and no additional recurrent fusions were detected (Supplementary Table S3).

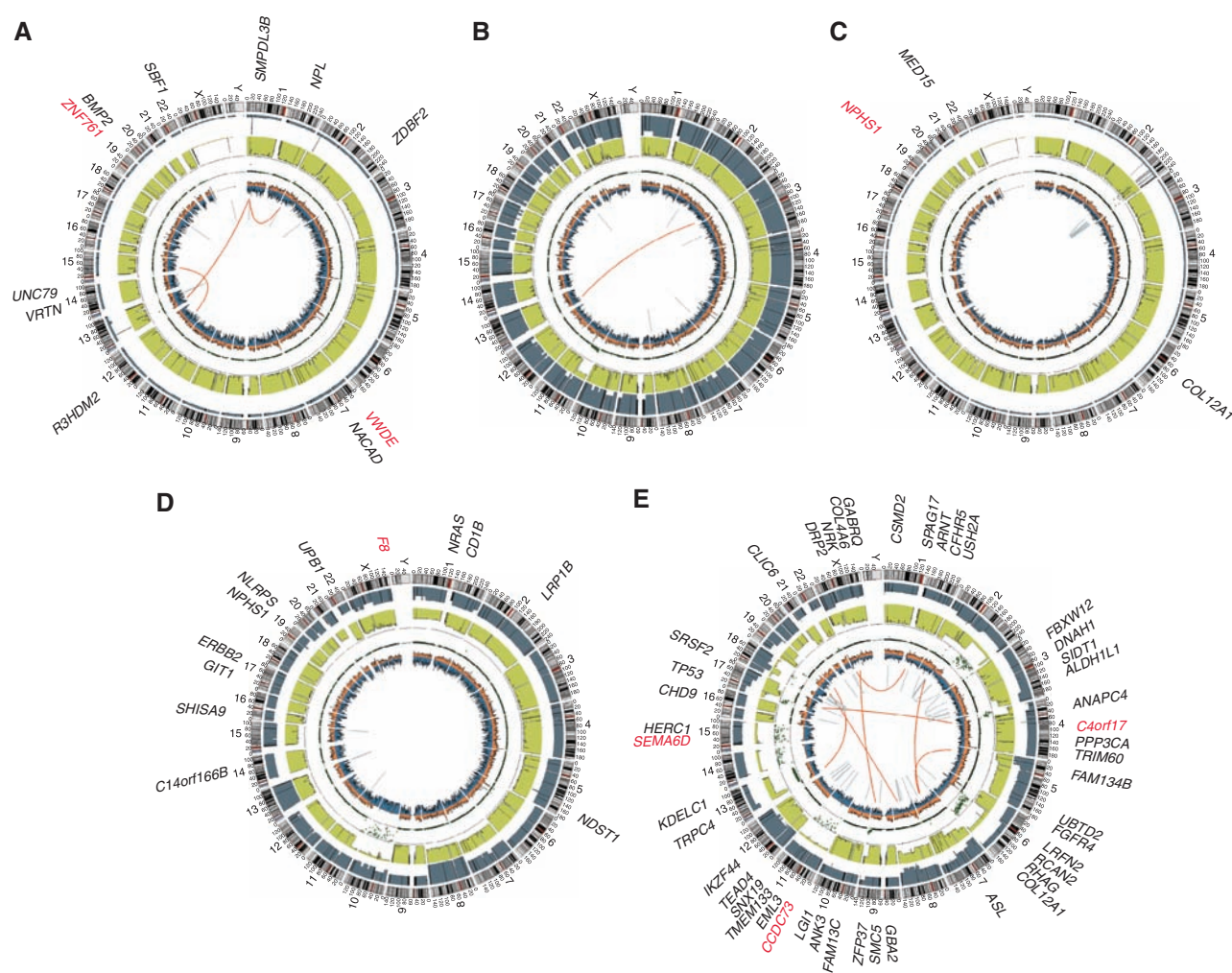
**Presence or Absence of a PAX Gene Fusion Defines Two Distinct Tumor Genotypes**

Of note, when PAX gene fusion-positive (PFP) tumors were compared with PAX gene fusion-negative (PFN) ones, we found a significantly increased mutation burden in the PFN population (Fig. 3A). On average, PFN tumors had significantly more verified somatic nonsynonymous mutations per tumor than PFP tumors (17.8 and 6.4, respectively;  $P = 2 \times 10^{-4}$ ). In contrast to the PFP samples, the PFN samples had an overall increase in aneuploidy ( $P = 1 \times 10^{-5}$ ; Fig. 3B). One remarkable PFP tumor (RMS224) from a 3-month-old patient had no protein-coding somatic alterations with the exception of the PAX3-FOXO1 fusion and copy-neutral LOH on chromosome 11p (Fig. 1B). Interestingly, both of the PFP and PFN genotypes seemed to have a distinct relationship between mutational frequency and age, with an increasing number of somatic mutations with older age of diagnosis and a steeper slope of curve in PFN tumors (Fig. 3C).

**Genes Recurrently Affected by Mutation in Rhabdomyosarcoma**

In total, we identified 542 somatic mutations (including missense, nonsense, splice site, and small insertions/deletions) altering 495 genes (40 recurrent) in the discovery set of 44 tumors (Supplementary Table S5); 58% of these alterations were predicted to be deleterious by Sorting Tolerant from





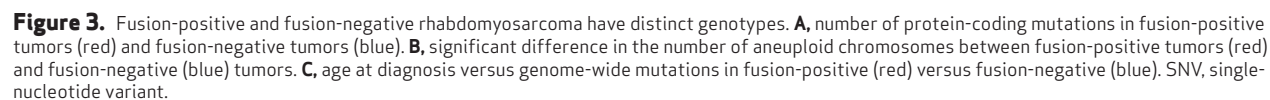
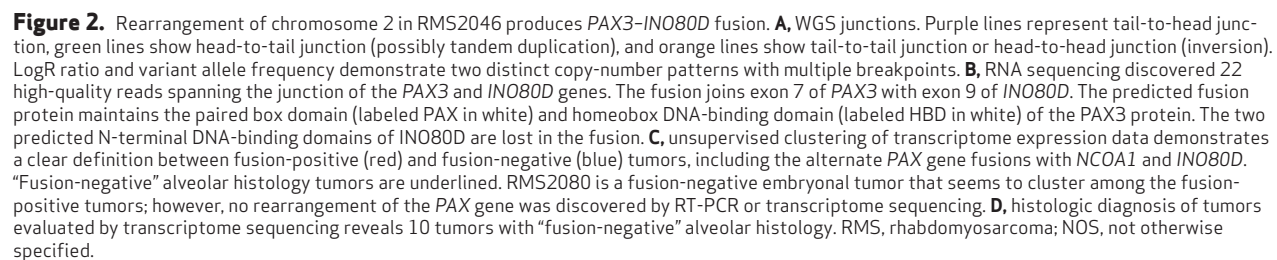
**Figure 1.** Circos plots of representative rhabdomyosarcoma tumors. Circos plot tracks representing verified somatic mutations, from outside circle; mutated genes missense mutations (black), nonsense and indel mutations (red); genomic location, genome wide copy-number alterations (gray), lesser allele frequency (green) LOH (dotted track), density of heterozygous SNPs (orange) homozygous SNPs (blue). Intrachromosomal rearrangements (inner circle gray) and interchromosomal rearrangements (inner circle red). **A**, NCI-40: a PAX7-FOXO1 translocation noting the associated high-level copy-number gain. This tumor also has high-level copy-number gain of MYCN on chromosome 2. **B**, RMS224: representative PAX3-FOXO1 fusion with no somatic point mutations. Note LOH on short arm of Chr11p. **C**, RMS2046: multiple rearrangements on chromosome 2 with corresponding junctions and copy-number changes. This rearrangement produces a novel gene fusion of PAX3-INO80D. **D**, RMS216: representative PFN rhabdomyosarcoma. Note relative increase in point mutations, including NRAS mutation on chromosome 1, and increase in aneuploidy, including gain of chromosome 8. Complete LOH on chromosome 10 and the short arm of chromosome 11. **E**, RMS2030: multiple genome-wide alterations in a tumor with TP53 mutation. Point mutation of FGFR4 on chromosome 5.

Intolerant (SIFT) analysis (17). These genes were selected for further verification and validation across the entire cohort and ranked using recurrence, background mutation rate, gene size, and nonsynonymous:synonymous ratio (Table 1). The list contained genes previously reported as altered in rhabdomyosarcoma, including *HRAS*, *KRAS*, *NRAS* (7), *FGFR4* (9), *PIK3CA*, and *NF1* (18), as well as genes not previously implicated in rhabdomyosarcoma such as *FBXW7* and *BCOR* (Fig. 4).

### Receptor Tyrosine Kinase/RAS/PIK3CA Mutations Predominately Affect PFN Tumors

Mutations affecting the receptor tyrosine kinase/RAS/PIK3CA pathway were the most common mutations observed in the study. Alterations in the RAS genes, *NRAS* (representative genome; Fig. 1D; PFN frequency, 11.7%), *KRAS* (PFN

frequency, 6.4%), and *HRAS* (PFN frequency, 4.3%) affected the oncogenic codons 12, 13, or 61 and were predominantly found in the embryonal rhabdomyosarcoma subtype as previously described (19); however, one tumor (RMS2051) with “fusion-negative” alveolar rhabdomyosarcoma histology carried an *NRAS* mutation. No RAS mutations were found in PFP tumors. Mutations in immediate effectors of RAS were also found, including alterations in the tumor suppressor *NF1* (PFN, 5.3% mutated; 17q11.2; LOH, 9%; Supplementary Fig. S2) and one tumor with an oncogenic mutation in *BRAF* at codon V600E (RMSS013). *PIK3CA* mutations (PFN, 7.4%) occurred at the known oncogenic codons Q546 or H1047, affecting the helical and the kinase domain, respectively. Interestingly, two samples (RMS2028 and RMS217) had concurrent mutation of *PIK3CA* and a RAS family gene. Despite a



**Table 1. Genes with significant frequency of somatic mutation across 147 rhabdomyosarcomas**

Gene	WGS	WES	Rhabdomyo- sarcoma recurrence	Rhabdomyo- sarcoma frequency	PFP recurrence	PFP frequency	PFN recurrence	PFN frequency	FDR
<i>NRAS</i>	4	7	11	7.5%	0	0	11	11.7%	5.10e-09
<i>FGFR4</i>	3	6	9	6.1%	0	0	9	9.6%	3.15e-12
<i>PIK3CA</i>	3	5	8	5.4%	1	1.9%	7	7.4%	5.86e-10
<i>BCOR</i>	3	5	8	5.4%	1	1.9%	7	7.4%	2.11e-08
<i>FBXW7</i>	3	4	7	4.8%	0	0	7	7.4%	2.11e-08
<i>KRAS</i>	3	3	6	4.1%	0	0	6	6.4%	5.51e-06
<i>TP53</i>	1	4	5	3.4%	0	0	5	5.3%	5.51e-06
<i>NF1</i>	3	2	5	3.4%	0	0	5	5.3%	2.06e-03
<i>HRAS</i>	2	2	4	2.7%	0	0	4	4.3%	1.12e-05

Abbreviation: FDR, false discovery rate.

predilection for embryonal rhabdomyosarcoma tumors (6 of 7), one fusion-positive alveolar rhabdomyosarcoma tumor (RMS244) also harbored a mutation in *PIK3CA*. Direct effectors of phosphoinositide 3-kinase (PI3K) were also found to be altered, including a predicted damaging mutation in *PIK3CD* (RMS2107) and homozygous deletion of *PTEN* (RMS2117). Across the whole population, several tyrosine kinase genes were found to be recurrently mutated, including *FGFR4* (PFP, 0%; PFN, 9.6%), *PDGFRA* (1.4%), and *ERBB2* (1.4%).

### Mutations of Cell-Cycle Genes and Other Key Pathways

Genes that control the cell cycle were also frequently mutated in the study population. *FBXW7*, an E3 ubiquitin ligase, was mutated in 7.4% of PFN tumors. All mutations in this gene occurred in the PFN subtype at conserved arginine residues (R387P, R441G, and R367P) within the WD40 repeat regions involved in substrate recognition. Mutations in the WNT signaling molecule *CTNNB1* are known drivers in colorectal cancer and medulloblastoma and have recently been described in rhabdomyosarcoma (8). In this study, we found three tumors (PFN 3%) with alterations at the known oncogenic codons S33 ( $n = 1$ ) and T41 ( $n = 2$ ), one of which occurred in a fusion-negative alveolar rhabdomyosarcoma tumor. Somatic mutation of *TP53* occurred exclusively in PFN tumors (PFN, 5.3%; representative genome; Fig. 1E), and 12% of all tumors had LOH of 17p13.1, which includes *TP53* (Supplementary Fig. S2). One patient (RMS212) was found to have a germline pathogenic mutation in *TP53* at R248 (Supplementary Fig. S3A–S3C). Other mitotic cell-cycle checkpoint genes were mutated at low frequencies, including *BUB1B* (1.4%), *FOXM1* (1.4%), *CCND1* (1%), and *CCND2* (1%). A notable finding was the recurrent alteration of *BCOR*, located on chromosome Xp11.4, in 7% of all rhabdomyosarcoma cases. Among these alterations in *BCOR*, nine were found in PFN tumors (seven mutations, two focal homozygous deletions), and one small indel was found in a PFP tumor (Table 1 and Supplementary Fig. S4).

### RNA Sequencing Highlights Expression of Candidate Oncogenes

To further enrich the analysis for potential oncogenes and targetable mutations, we performed mutational analysis of the RNAseq data of 80 tumors (29 PFP and 51 PFN) to determine which of the somatic mutations found at the DNA level were also expressed in the transcriptome. Fifty-eight percent of the verified somatic mutations discovered at the DNA level had evidence of RNA expression (Supplementary Fig. S5A and S5B). Each tumor harbored a median of five expressed somatic mutations (range, 0–26; PFN median, nine mutations; PFP median, 2.5 mutations; Supplementary Table S6). Thirty-three genes were found to recurrently harbor expressed mutations, including *PTPN11* (0 PFP vs. 2 PFN; ref. 20), the DNA repair gene *ATM* (2.5%; 1 PFP vs. 1 PFN), BRCA1-interacting protein *ZNF350* (2.5%; 1 PFP vs. 1 PFN), and MYCN-interacting protein *TRPC4AP* (2.5%; 0 PFP vs. 2 PFN). In addition, we discovered expressed singleton mutations in *FOXO1* and *ARID1A* (Fig. 5A) not previously observed in rhabdomyosarcoma. By Gene Ontology, the expressed mutations were markedly enriched for genes involved in cell cycle ( $P = 2e-6$ ), protein phosphorylation ( $P = 6.9e-5$ ), DNA damage ( $P = 1.3e-4$ ), muscle cell differentiation ( $P = 3.3e-4$ ), regulation of mitogen-activated protein kinase (MAPK) activity ( $P = 3.3e-4$ ), chromatin modification ( $P = 9e-4$ ), and induction of apoptosis ( $P = 2.8e-3$ ; Supplementary Table S7). Many of the tumors seemed to accumulate multiple genetic hits within these pathways (Fig. 5B).

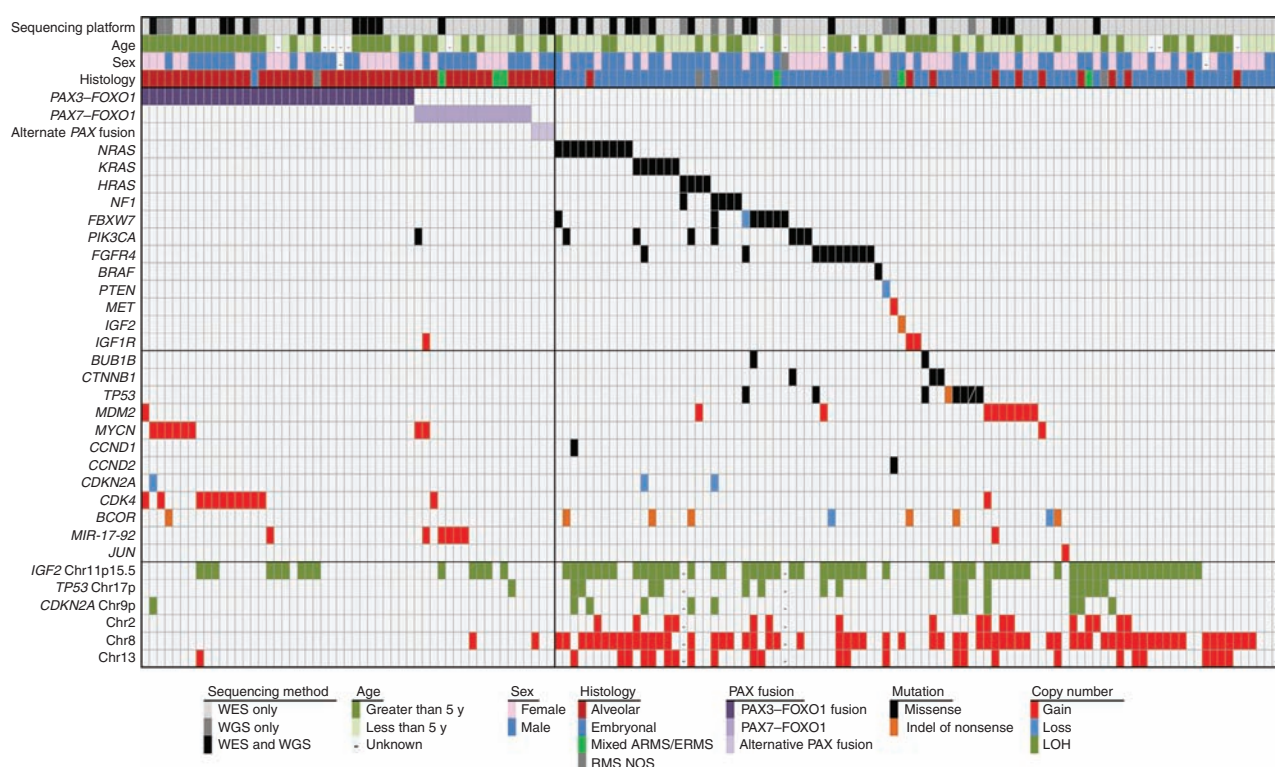
### Copy-Number Alterations

To evaluate somatic copy-number alterations (CNA) important in rhabdomyosarcoma, high-resolution (2.5 or 5 million) SNP array analyses were performed on all tumors, and recurrent focal amplifications and deletions were analyzed by frequency in the study population (Supplementary Fig. S6). LOH of 11p15.5 was found in 50% (16 PFP vs. 59 PFN) of the surveyed tumors. The minimum common region of overlap encompassed a region of 11p15.5 (Supplementary



## RESEARCH ARTICLE

Shern et al.



**Figure 4.** The genomic landscape of pediatric rhabdomyosarcoma (RMS) highlighting candidate alterations. Demographic characteristics, histologic subtypes, and selected genes with copy-number alterations or somatic mutations across 147 rhabdomyosarcoma cases. Unique sample identifier and sequencing platform. Sex, males in blue, females in pink. Age, years at diagnosis divided into fewer than 5 years and greater than 5 years. Histologic diagnosis, red, alveolar rhabdomyosarcoma (ARMS); blue, embryonal rhabdomyosarcoma (ERMS) including spindle and botryoid subtypes; gray, rhabdomyosarcoma not otherwise specified (NOS). Mixed alveolar and embryonal histology in green. Copy-number gains and losses for selected genes. Blue, losses; red, gains; green, LOH. Selected genes with somatic mutations. Purple, fusion protein; black, missense; orange, nonsense/splice site/indel mutations.

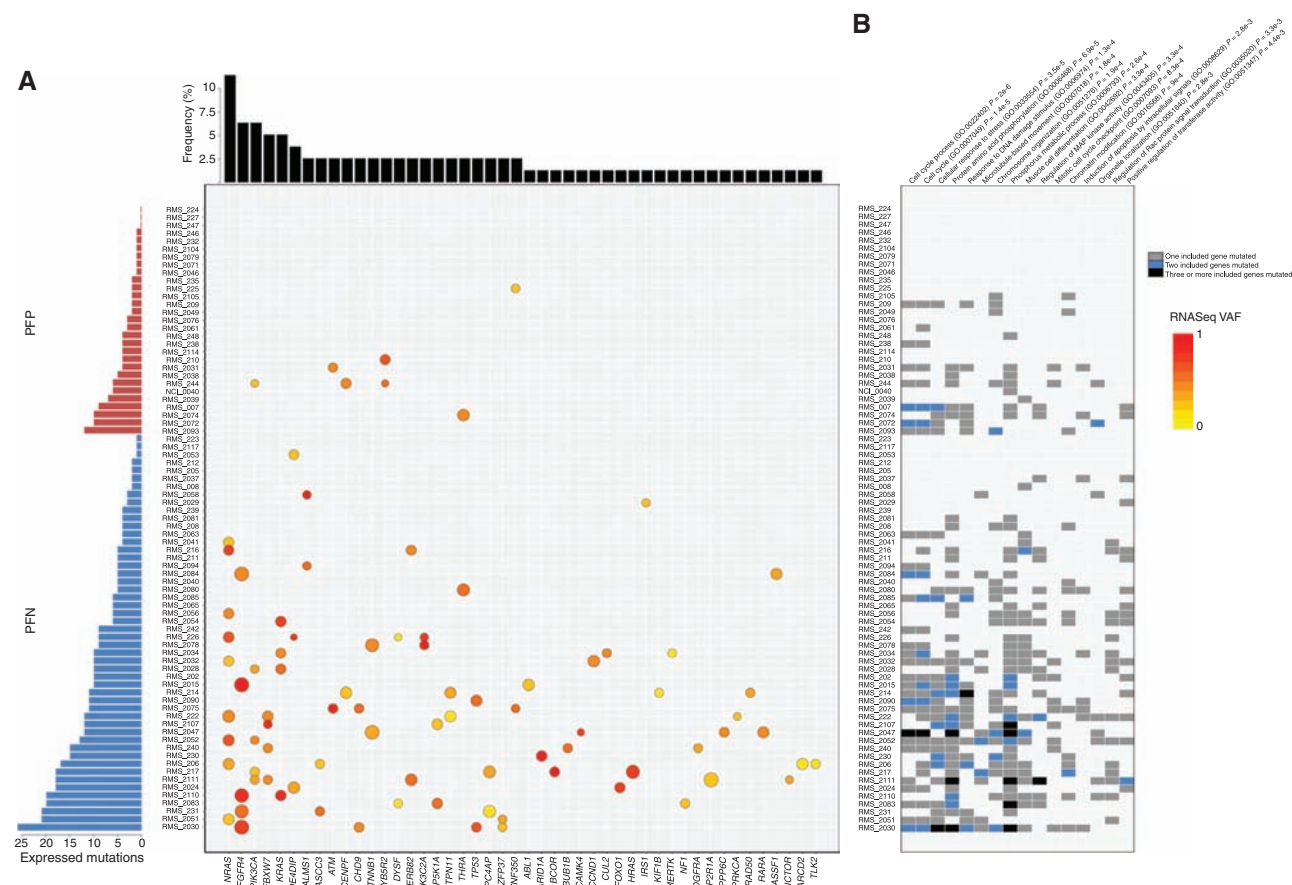
Fig. S7) that includes the paternally imprinted gene *IGF2*. Further evidence of insulin receptor signaling alterations in rhabdomyosarcoma were observed with focal amplification of *IGF1R* in 2.7% (1 PFP vs. 3 PFN) of cases (Supplementary Fig. S8A; ref. 21) and one somatic indel in the 3'-untranslated region of *IGF2* (RMS2037; Supplementary Table S5). Consistent with previous reports, 9.7% of the tumors displayed amplification of chromosomal region 12q13-q14, which has been shown to be associated with worse overall survival in rhabdomyosarcoma independent of gene fusion status (22). The 12q13-q14 amplicon was found predominantly in PFP tumors (12 PFP vs. 1 PFN). The minimum amplicon size (Supplementary Fig. S8B) included 25 genes, including the cyclin-dependent kinase gene *CDK4*. Recurrent focal amplification of 12q15 (9%; Supplementary Fig. S8C) that encompassed the genes *FRS2* and *MDM2* occurred predominantly in PFN tumors (9 PFN vs. 1 PFP). Amplification of 2p24 involving *MYCN* (5%) occurred predominantly in PFP tumors (8 PFP vs. 1 PFN; Supplementary Fig. S8D); amplification of the *PAX7-FOXO1* fusion gene occurred in 12/15 *PAX7-FOXO1* tumors (Supplementary Fig. S8E), and amplification of 13q31-32 including the *MIR17HG* locus occurred exclusively in PFP tumors (4.5%; Supplementary Fig. S8F). Homozygous deletion of the tumor suppressor *CDKN2A* was found in 3% of samples, and LOH at this locus

(9p21.3) occurred in 9% (1 PFP vs. 13 PFN) of the study population. This allelic loss rate was lower than the previously reported frequency of 25% (15). As previously described (23), recurrent gain of chromosome 8 was seen in 46% of the PFN population. Other chromosome level events included recurrent gains of chromosomes 2, 7, 11, and 13 and the recurrent loss of chromosomes 1p, 9, and 16 (Supplementary Fig. S9).

### Pathway Analysis Integrating Mutations, Copy-Number Changes, and Structural Variations Implicates Alteration of FGFR Signaling

To identify dysregulated pathways relevant to rhabdomyosarcoma pathology, analysis incorporating structural variations, copy-number changes, and somatic mutations found in the WGS discovery set was performed. Using the 2,119 genes found to be somatically altered in the discovery cohort (Supplementary Table S8), Reactome (24) overrepresentation analysis indicated that fibroblast growth factor receptor (FGFR) signaling was the most significantly altered pathway ( $P = 4.6 \times 10^{-5}$ ), with 29 of 112 candidate genes represented. Remarkably, mutations in this pathway (Fig. 6A) were found in 88% of PFN samples (22 of 25 tumors) that were analyzed by WGS. When examined separately, the genes altered in PFP tumors (435 of 2,119) were not significantly enriched in any canonical pathways.





**Figure 5.** Expressed mutations in 80 rhabdomyosarcoma tumors. **A**, candidate somatic alterations found to be expressed in WTS and the discovered genes ranked by frequency. Top, the number of expressed mutations by sample; blue, PFN; red, PFP. The color (yellow to red) of the mark represents the variant allele frequency (VAF) with many mutations appearing to favor the mutant allele. The size of the circle is proportional to the fragments per kilobase of transcript per million mapped reads (FPKM). **B**, Gene Ontology analysis of the expressed mutations reveals multiple alterations of cell cycle, cellular response to stress, protein amino acid phosphorylation, response to DNA damage stimulus, microtubule-based movement, chromosome organization, muscle cell differentiation, regulation of MAPK activity, mitotic cell-cycle checkpoint, chromatin modification, induction of apoptosis by intracellular signals, organelle localization, regulation of RAC protein signal transduction, and regulation of transferase activity. Many tumors seem to accumulate multiple mutations in the same pathway (blue = 2; black = 3 or more).

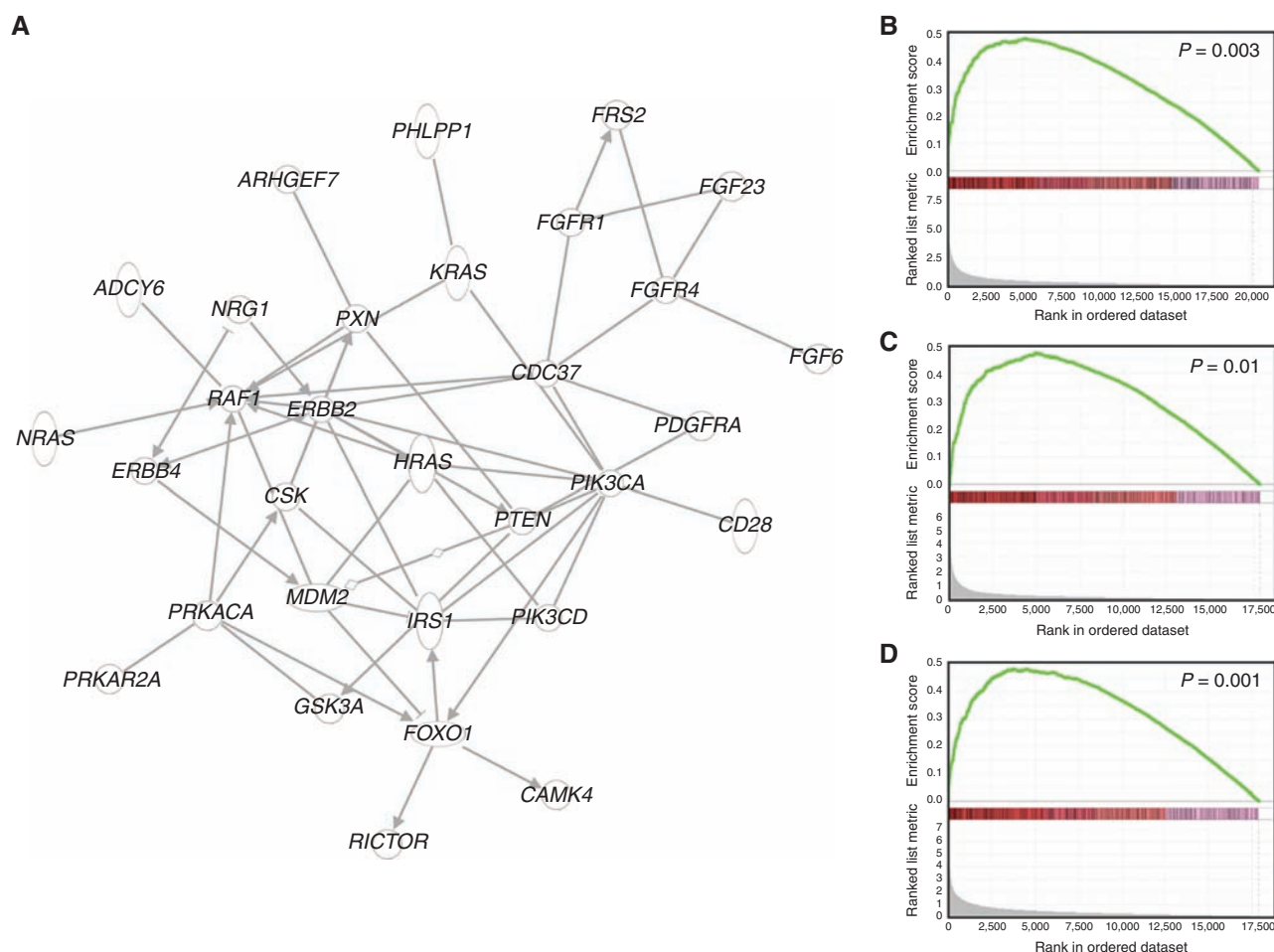
### PAX3-FOXO1 Model System Reveals Alteration of a Common Genetic Axis in Fusion-Positive and Fusion-Negative Tumors

Of note, several genes found altered in PFN tumors, including *MYO1D1*, *MET*, *CNR1*, and *FGFR4*, are known downstream targets of PAX3 and PAX3-FOXO1 ( $P = 1.54 \times 10^{-3}$ ; ref. 25), leading us to hypothesize that mutations in PFN tumors may be enriched for genes regulated downstream of the PAX fusion proteins. To experimentally test this hypothesis, we constructed a human fibroblast cell line stably expressing PAX3-FOXO1 (cell line 7250\_Pf) and used expression arrays to compare it with the isogenic control (26). This analysis identified 444 genes that had greater than 4-fold change when PAX3-FOXO1 was expressed. Top upregulated genes included multiple genes that were found to be mutated in PFN tumors, such as *FGFR4*, *CCND2*, and *IGF2* (Supplementary Table S9). As confirmation of the model system, these differentially expressed genes were also overrepresented in

the 76 genes recently reported as PAX3-FOXO1 targets ( $P = 1.7 \times 10^{-3}$ ) by using chromatin immunoprecipitation sequencing (ChIPseq; ref. 27). Remarkably, the 2,119 somatically altered genes identified in our WGS samples were significantly enriched in the differentially expressed genes modulated by PAX3-FOXO1 in the fibroblast cell line experiment using gene set enrichment analysis (GSEA;  $P = 3 \times 10^{-3}$ ; Fig. 6B). The observed enrichment was more prominent for the mutated genes from PFN tumors ( $P = 7 \times 10^{-3}$ ) than those from PFP samples ( $P = 0.08$ ). To further validate this hypothesis, we repeated GSEA analyses using published data derived from a transgenic mouse model expressing the *PAX3-FOXO1* fusion gene in the developing forelimb or somites (28). Consistent with our *in vitro* results, mutated genes in the PFN tumors were significantly enriched in both the forelimb and somite datasets ( $P = 0.001$  and  $0.01$ ; Fig. 6C and D), whereas there was no enrichment for those in the PFP tumors ( $P = 0.159$  and  $0.543$ ). A set of 116 common genes, including *FGFR4*, was found in the leading edge of all

## RESEARCH ARTICLE

Shern et al.



**Figure 6.** **A**, gene interaction map of Reactome pathway analysis that discovers alteration of FGFR signaling as the most altered pathway. Twenty-two of 25 fusion-negative tumors alter at least one gene in the pathway. **B**, GSEA enrichment plot of altered genes in fusion-negative rhabdomyosarcoma tumors versus altered genes in the PAX3-FOXO1-expressing model cell line (7250\_Pf) with enrichment scores plotted for each gene moving down the ranked list of genes. Genes altered in the 7250\_Pf cell line show significant enrichment in the fusion-negative tumors (FDR  $q$ -value = 0.004). GSEA enrichment plot of the altered genes in published mouse models of PAX3-FOXO1 in **C**, somite cells (FDR  $q$ -value = 0.009) and **D**, forelimb cells (FDR  $q$ -value 0.0009). FDR, false discovery rate.

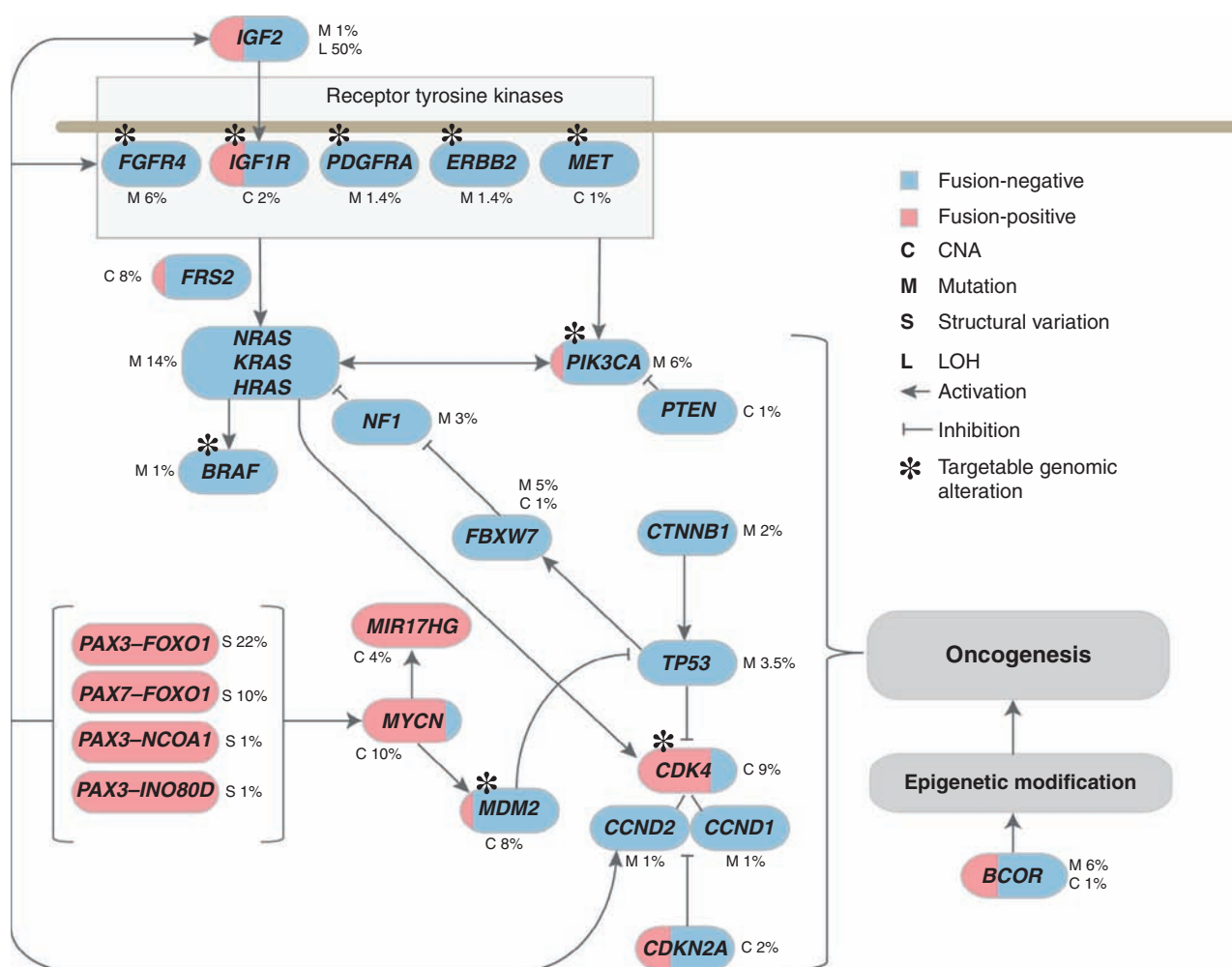
three PAX-fusion model systems in the GSEA analyses (Supplementary Table S10).

## DISCUSSION

To our knowledge, this study represents the most comprehensive characterization reported to date for the genomic alterations that underlie rhabdomyosarcoma. We found that subcategorization by the presence or absence of a *PAX* gene fusion more accurately captures the true genomic landscape and biology of rhabdomyosarcoma than the traditional alveolar rhabdomyosarcoma/embryonal rhabdomyosarcoma histologic distinction. This finding is consistent with the clinical observation that the presence or absence of a *PAX3/7-FOXO1* gene fusion is a crucial prognostic indicator in this disease (29, 30) and that fusion-negative alveolar rhabdomyosarcoma seems to mimic the clinical course of embryonal rhabdomyosarcoma in the majority of patients (23). Despite this, our findings indicate that there is a subpopulation of fusion-negative alveolar

rhabdomyosarcoma that harbor a rearrangement of the *PAX3* gene with a cryptic partner, a finding that may have important clinical ramifications for the proper therapeutic stratification of patients.

Overall, the low somatic mutation rate that we observed is consistent with large sequencing efforts of other pediatric solid tumors and presents an enormous correlative and clinical challenge (31–34). In rhabdomyosarcoma, this was particularly evident in tumors that harbored a translocation oncogene (0.1 protein coding changes per megabase). This finding underscores the importance of the *PAX* gene fusion as the dominant driver in this subtype, which through its transcriptional reprogramming alters a host of downstream targets. However, it is important to note that multiple genetic model systems have shown that *PAX3-FOXO1* by itself cannot cause rhabdomyosarcoma and that a coexisting genetic lesion is necessary (35, 36). Experimentally validated cooperating lesions in a mouse model of alveolar rhabdomyosarcoma include *TP53* and *INK4a/ARF* loss (37).



**Figure 7.** Model pathway altered in rhabdomyosarcoma. Genes colored red are found in fusion-positive tumors, whereas genes colored blue are found in tumors without a PAX gene fusion. Alterations and their frequency in the population include mutations and small indels (M), copy number deletions and amplifications (C), or structural variations (S) that affect the gene.

Our data demonstrate that most commonly the cooperating event is due to genetic amplification (such as *MYCN*, *CDK4*, and *MIR-17-92*) or deletion (*CDKN2A*, LOH of Chr11p15.5), and only in rare cases can an additional candidate somatic driver mutation be nominated. In contrast, fusion-negative tumors seem to have accumulated a higher degree of aneuploidy and mutational burden at the time of clinical presentation.

Despite the relatively low mutation rate, rhabdomyosarcoma tumors do harbor a significant array of alterations, including chromosomal rearrangement, amplification, deletion, and mutation of recurrent drivers and novel candidate therapeutic targets. Many of the genetic alterations identified in this study, including *FGFR4*, *IGF1R*, *PDGFRA*, *ERBB2/4*, *MET*, *MDM2*, *CDK4*, and *PIK3CA*, are targeted by approved or late-stage therapeutics that could immediately inform clinical trials in rhabdomyosarcoma (Fig. 7). In this study, we found that the RAS pathway (including *FGFR4*, *RAS*, *NF1*, and *PIK3CA*) is mutationally activated in at least 45% of PFN tumors. Although directly targeting constitutively active RAS

remains challenging, the recent success of the MAP-ERK kinase 1/2 (MEK1/2) inhibitor trametinib in melanomas with mutated *NRAS* demonstrates the utility of inhibiting the effector pathways altered by the mutation (38). Early preclinical evidence has found efficacy of this method in rhabdomyosarcoma (39), and further efforts to precisely dissect the RAS effector pathways that are critical in rhabdomyosarcoma are currently under way.

A novel finding in this study is the discovery of recurrent mutations in *BCOR* affecting 7.4% of PFN tumors. *BCOR* is a transcriptional repressor that has been shown to interact with both class I and II histone deacetylases (40), and somatic mutations in *BCOR* have been described in other pediatric tumors including acute myeloid leukemia (32), retinoblastoma (33), and medulloblastoma (34). Our discovery of its recurrent alteration in rhabdomyosarcoma reinforces this chromatin modifier (41) as a potential therapeutic target. Further functional validation of the discovered mutations in *BCOR*, *ARID1A*, *ZNF350*, *TRPC4AP*, and others may provide targets for novel treatments in patients with rhabdomyosarcoma.



## RESEARCH ARTICLE

Shern et al.

Incorporation of the discovered genes into prospective, well-annotated clinical trials will be crucial in extending these findings' utility as therapeutic biomarkers.

Despite the challenges of low frequency of recurrence, the genetic study of pediatric cancer provides remarkable insight into the likely drivers of tumorigenesis by reducing the background of passenger mutations that naturally occur during aging. The observation that PFP and PFN genotypes seem to have a distinct relationship between mutational frequency and age, with a steeper slope in the PFN tumors, may have interesting implications. This finding suggests that PFN tumors require the accumulation of mutations before presentation, whereas malignant transformation of PFP tumors requires few somatic alterations beyond the occurrence of the fusion gene. This observation may also be due to differences in the respective tumor types' cell of origin, proliferation, and apoptotic rate, or an underlying DNA repair deficit. Our observation that 58% of the verified somatic mutations discovered at the DNA level had evidence of RNA expression is a higher proportion than the 36% rate observed in adult cancers such as breast cancer (42) or lymphoma (43), and may reflect an enrichment of driver mutations or the presence of fewer accumulated passenger mutations in these pediatric patients. In many cases, the expression of a mutated gene seems to be relatively increased and favors the variant allele. This finding, at least in theory, provides tractable genetic targets against which therapies could be developed.

Finally, our integrative analysis demonstrates that despite remarkable genetic and molecular heterogeneity, rhabdomyosarcoma tumors seem to hijack a common receptor tyrosine kinase/*RAS/PIK3CA* genetic axis. This occurs through two alternative mechanisms—either by rearrangement of a *PAX* gene or accumulation of mutations in genes that are downstream targets of the *PAX* fusion protein. Evidence for alteration of this common genetic axis can be found in 93% (41 of 44) of the tumors surveyed by WGS and seems to hinge on the fibroblast and insulin receptor pathways. These observations are consistent with previous proteomic studies of rhabdomyosarcoma (44, 45) and warrant continued biologic investigation and pharmacologic targeting of this axis as crucial to expanding the available therapeutic options. In conclusion, we report here the most comprehensive analysis of the genomic landscape of rhabdomyosarcoma to date. Our discoveries provide a rational framework for new avenues of translational research, including molecular subclassification and developing novel therapeutic strategies for children suffering from rhabdomyosarcoma.

## METHODS

### Sample Selection

All patient sample collection was approved by the Institutional Review Board of the participating facility. Samples were assembled from collections at the Pediatric Oncology Branch of the National Cancer Institute, Children's Oncology Group, the Tumour Bank at The Children's Hospital at Westmead (Westmead, NSW, Australia), and the Department of Oncology, St. Joan de Deu De Barcelona (Barcelona, Spain). All tumors were collected at initial diagnosis

and before any therapy, with the exception of samples NCI0040 and NCI0080 that were collected at relapse. Samples were deidentified and histologic diagnosis and clinical information were compiled. The selected tumors were >70% tumor:normal tissue on histology review when available. Quality control genotyping for the whole-genome samples was performed to ensure the match of tumor-normal pairs.

### Nucleic Acid Extraction and Whole-Genome Amplification

DNA was isolated from 10 to 25 mg of tumor or 1 mL of whole blood using the QIAamp DNA Mini Kits (Qiagen) or Agencourt Genfind v2 Kits (Beckman Coulter), respectively, according to the manufacturer's protocol. For WGS, approximately 6  $\mu$ g (range, 5–10  $\mu$ g) of native genomic DNA was sequenced according to the Complete Genomics method (46). Whole genome-amplified genomic DNA using high-fidelity Phi29 polymerase (Qiagen REPLI-g) was used for the whole-exome validation cohort according to the manufacturer's protocol (Qiagen). Quantification of DNA was performed using the Quanti-iT DNA assay (Life Technologies). Each DNA sample was examined by electrophoresis on a 1% agarose gel to ensure high quality. RNA extraction was accomplished with the Qiagen RNeasy Micro Kits according to the manufacturer's protocol (Qiagen).

### Calculation of Background Mutation Rate

The background mutation rate was calculated using the method described by Zhang and colleagues (47). Briefly, the background mutation rate is the silent mutation rate in coding regions adjusted by silent-to-nonsilent ratio [estimated to be 0.350 by The Cancer Genome Atlas (TCGA) Consortium] across the coding regions.

### Small-Variant Discovery for WGS

Small variants, including single-nucleotide variants and indels, were called using cgatools (<http://cgatools.sourceforge.net/docs/1.6.0/>) in build hg19. Somatic variants were determined first by comparison of the tumor with matched leukocyte normal DNA. To remove artifacts specific to the sequencing platform, we eliminated any somatic variants also found in normal subjects other than patients with rhabdomyosarcoma [50 in-house normal samples and 69 Complete Genomics samples (<http://www.completegenomics.com/public-data/69-Genomes/>)]. The Somatic Score (refs. 48–50; [http://info.completegenomics.com/rs/completegenomics/images/Cancer\\_Application\\_Note](http://info.completegenomics.com/rs/completegenomics/images/Cancer_Application_Note)) is based on a Bayesian model and takes account of read depth, base call quality, mapping/alignment probabilities, and measured priors on sequencing error rate for both the germline variants and the tumor variants. Using an independent platform (SOLiD exome sequencing) for verification of somatic variants from the WGS, we found an optimal balance between sensitivity and specificity by selecting variants with somatic score  $\geq 0$  (Supplementary Fig. S10A and S10B). Finally, small variants within regions that have significant similarity to other regions in the genome, taken from the "Self Chain" track of UCSC genome browser (<http://genome.ucsc.edu/cgi-bin/hgTracks?org=human>), were removed, as they are likely due to mapping errors.

The somatic variants were then annotated using ANNOVAR (51), which details the synonymous/nonsynonymous nature of the alteration, the corresponding amino acid alteration, as well as the presence or absence of the alteration in the Single Nucleotide Polymorphism Database (dbSNP) 135 and 1000 Genome Project. SIFT ([http://sift.jcvi.org/www/SIFT\\_chr\\_coords\\_submit.html](http://sift.jcvi.org/www/SIFT_chr_coords_submit.html)) and Polyphen (<http://genetics.bwh.harvard.edu/pph2/bgi.shtml>) scores were used to determine the potential impact of an SNP variant. Oncotator (<http://www.broadinstitute.org/oncotator/>) was used to add cancer-specific annotations from the Catalogue of Somatic Mutations in Cancer (COSMIC) and TCGA.

### Verification of WGS-Predicted Somatic Mutations

Verification of small variants with somatic score no less than 0 predicted by WGS was accomplished by comparing overlapping exome sequencing that was carried out on 30 of 44 tumor samples. Each somatic position was examined for the identical change as well as coverage in the exome sequencing data. Verification of the mutation was called when greater than 3 exome reads supported the WGS read. Additional site verification was performed with barcoded DNA libraries made from the 44 WGS tumors using a designed Custom AmpliSeq Cancer Panel and AmpliSeq Library Kit 2.0 with sequencing on the Ion Torrent Personal Genome Machine. Multiplex PCR library preparation, emulsion PCR (ePCR) template preparation, and semiconductor sequencing were performed according to the manufacturer's protocol. Sequencing generated 868 Mb high-quality bases with average amplicon coverage of 206 $\times$ . Using this method, sensitivity is calculated at 84% (assuming the small variants reported by Complete Genomics WGS includes all the true positive variants) and a specificity at 93%. Additional verification of mutations reported in Fig. 1 and Table 1 was accomplished by PCR amplification of genomic DNA using a uniform annealing temperature of 65°C followed by standard Sanger sequencing and analysis using Sequencher 4.10 software (Gene Codes Corporation).

### Copy-Number Discovery from WGS

The Complete Genomics copy-number segments (based on 2-kb window) profile was used to call amplifications ( $\geq 5$  copies) and homozygous deletions. CNAs were divided into two groups: (i) focal amplification or deletions less than one arm in length; and (ii) whole-arm or whole-chromosome events.

### Junction Discovery

On the basis of the high-confidence junction reports for the tumor sample and the paired germline sample, we called the somatic junctions as those present only in tumor samples. Somatic junctions that are present in other normal samples (50 in-house germline samples and 69 Complete Genomics baseline germline samples) were removed to reduce systematic artifact. Comparison of the predicted somatic junctions with the corresponding SNP array copy-number data shows that 90% of the junctions had changes in copy-number state or allelic ratio at the predicted break point.

### Circos Plots

Circos plots were generated for each sample using the Circos plotting software provided by Complete Genomics (<http://www.completegenomics.com/analysis-tools/cgatools>), with in-house customized modifications.

### RT-PCR of PAX-FOXO Gene Fusion

We determined the *PAX3-FOXO1* or *PAX7-FOXO1* fusion status using RT-PCR of tumor RNA, using specific oligonucleotide primers according to the published method (52).

### RNA Sequencing

PolyA selected RNA libraries were prepared for RNA sequencing on Illumina HiSeq2000 using TruSeq v3 chemistry according to the manufacturer's protocol (Illumina). Hundred bases-long paired-end reads were assessed for quality and reads were mapped using CASAVA (Illumina). The generated FASTQ files were used as input for TopHat2 (53). Using SAMtools (<http://samtools.sourceforge.net/>), the produced BAM files were compared with the sites found somatically mutated in DNA, and total coverage and variant allele frequency (VAF) were calculated. Expressed fusion transcripts were detected by tophat-fusion 0.1.0 (54) and deFuse 0.4.3 (55) with hg19 human genome assembly.

### RNAseq Expression Analysis and Unsupervised Clustering

Cufflinks (<http://cufflinks.cbc.umd.edu/>; ref. 56) was used to assemble and estimate the relative abundances of transcripts mapped with TopHat2 at the gene and transcript level (FPKM). FPKM values were log<sub>2</sub> transformed. Samples were clustered on the basis of Ward's algorithm based on Euclidean distance.

### SOLiD Exome Sequencing and Data Analysis

We constructed sequencing libraries and performed target enrichment by using the Agilent SureSelect Human All Exon Kits designed to target 37.8-Mb regions of all human exons according to the manufacturer's instructions (Agilent). The PCR-amplified libraries were sequenced on SOLiD 4 systems using the 50  $\times$  35 bp paired-end sequencing protocol (Applied Biosystems). Sequencing was used to evaluate 120 tumor-normal pairs for coding sequence alterations. An average of 3.3 Gb of nonredundant sequence was mapped on-target per sample to hg19 using BFAST version 0.7.0a (57). Duplicates were removed using Picard (<http://picard.sourceforge.net/command-line-overview.shtml>); normal-tumor BAM files were used as input for GATK version 2.1-11 (<http://www.broadinstitute.org/gatk/>; refs. 58–60). Local realignment and base quality recalibration were performed using default parameters. SNPs and indels were called using GATK UnifiedGenotyper ([http://www.broadinstitute.org/gatk/gatkdocs/org\\_broadinstitute\\_sting\\_gatk\\_walkers\\_genotyper\\_UnifiedGenotyper.html](http://www.broadinstitute.org/gatk/gatkdocs/org_broadinstitute_sting_gatk_walkers_genotyper_UnifiedGenotyper.html)). Variants that passed quality score greater than 50 coverage in tumor and normal greater than 10, VAF in the tumor greater than 15% and VAF in normal of 0% were further annotated with ANNOVAR, SIFT, PPH2, and COSMIC.

### Illumina Exome Method

Tumor and tumor DNA (100 ng) underwent shearing, end repair, phosphorylation, and ligation to barcoded sequencing adaptors. The ligated DNA was size-selected for fragments between 200 and 350 bp. This prepared DNA underwent exome capture using SureSelect v2 Exome bait (Agilent). Captured DNA was multiplexed and sequenced on Illumina HiSeq flow cells. Exome analysis was performed using Broad Institute pipelines (61, 62). MuTect and MutSig algorithms were used to call somatic mutations and determine statistical significance, respectively.

### Comparison of PFP Mutation Rate with PFN Mutation Rate

We hypothesized that there are fewer somatic nonsynonymous mutations in PFP than in PFN rhabdomyosarcoma (Fig. 3). To test this hypothesis, we compared the number of verified somatic nonsynonymous mutations in PFP with that in patients with PFN rhabdomyosarcoma. A random permutation test (RPT) was performed by permuting the group label of the patients, to avoid any assumption of the unknown distribution of the number of mutations as well as the bias brought by the assumption. The summary statistic is defined as the between-group variability divided by the within-group variability, to measure the difference between the two patient groups while considering the variability within each patient group.

### Quantifying the Relation between Mutational Frequency and Age

We observed that the PFP and PFN rhabdomyosarcoma have distinct but consistent relationships between mutational frequency and age (Fig. 3C). Therefore, we applied linear regression to model the relationship between the number of WGS single-nucleotide variants (dependent variable) and the corresponding age at diagnosis (exploratory variable) for patients with PFP (16 patients with age information) and PFN (24 patients with age

## RESEARCH ARTICLE

Shern et al.

information), respectively. The goodness of fit of each linear regression model is measured via the *t* test on the Pearson correlation coefficient between the mutation number reported by WGS and that predicted from the linear regression model (one-tailed test due to the preknown positive correlation).

### Method of Determining Statistically Significant Genes

The 621 genes harboring small variants with somatic score no less than 0 in the WGS study were applied to the exome validation samples (103 total tumor-normal pairs—90 SOLiD/13 Illumina). Included in the recurrence calculation is any small variant that was called using the aforementioned SOLiD or Illumina GATK analysis pipeline. Total recurrence was then calculated as the number of mutations in a gene in the WGS data in addition to the WES data. Ranking of this gene list was done by using the binomial method reported by Wei and colleagues (63) to calculate the significance of a gene mutation. This method considers the recurrence of mutation in the observed gene, the length of the gene coding region, and the background mutation rate as well as the synonymous:nonsynonymous ratio. Significant genes were selected on the basis of false discovery rate (FDR)  $\leq 0.05$  (64).

### SNP Array

Illumina Omni 2.5M (97 paired plus 30 unmatched tumors) or 5M (10 paired samples) were performed according to the standard procedure from the manufacturer (Illumina) at the National Cancer Institute (NCI), Cancer Genomics Research Laboratory. When available, matched normal:tumor paired arrays were analyzed. For copy-number analysis, the raw data were processed and normalized in Illumina GenomeStudio ([http://www.illumina.com/Documents/products/technote/infinitum\\_genotyping\\_data\\_analysis.pdf](http://www.illumina.com/Documents/products/technote/infinitum_genotyping_data_analysis.pdf)). Final reports were exported and imported into Nexus BioDiscovery (<http://www.biodiscovery.com/downloads/pdfs/SimplifyingDataInterpretationWithNexusCopyNumber.pdf>) software in paired mode. In Nexus, the data were corrected for GC content and segmented by using SNP-FASST2. Frequency across the whole population and according to the fusion status was analyzed using the Significance Testing for Aberrant Copy number (STAC) algorithm (65). High copy amplicons were plotted using the row probe level log relative ratios.

### Pathway Analysis

Reactome pathway analysis (<http://www.reactome.org/>) was performed as previously described (24). Overrepresentation analysis of the somatically altered genes was performed. Fisher exact test was used to calculate a *P* value determining the probability that the association between the genes in the dataset and the observed pathway is explained by chance alone. Further gene interaction analysis was performed through the use of Ingenuity Pathway Analysis (IPA; Ingenuity Systems; [www.ingenuity.com](http://www.ingenuity.com)). Functional analysis was performed in which the biologic functions most significant to the dataset were extracted. A right-tailed Fisher exact test was used to calculate a *P* value determining the probability that each biologic function assigned to that dataset is due to chance alone.

### Association between the Genes Altered in PFN Rhabdomyosarcoma and the PAX-FOXO1-Binding Genes

We observed that many genes frequently altered in fusion-negative rhabdomyosarcoma tumors are PAX-FOXO1-binding genes. To test whether the overlap between these two gene groups is by chance or not, we compared the genes with somatic mutations, copy-number amplification, copy-number homozygous-deletion, or structural variants (1,957 genes reported by WGS) in 25 fusion-negative rhabdomyosarcoma tumors to the genes recently reported as significantly altered in a chromatin precipitation identification of PAX3-FOXO1-

binding sites (76 genes reported in Cao; ref. 27). Fisher exact test was performed to assess the association between these two gene groups and  $P = 4.5 \times 10^{-3}$ , rejecting the null hypothesis (the significance threshold is set as 0.05). This result indicates that the genes altered in patients with PFN rhabdomyosarcoma are significantly associated with the PAX-FOXO1-binding genes.

### 7250\_PF Cell Line

The stably transfected cell line expressing PAX3-FOXO1 was constructed as previously described from the parent cell line CRL7250 (American Type Culture Collection; refs. 26, 66). The cell lines were validated by the Division of Cancer Epidemiology and Genetics, National Cancer Institute, by using short-tandem repeat DNA fingerprinting. The expected expression of the PAX3-FOXO1 fusion oncogene was evaluated with RT-PCR (Supplementary Fig. S11). All cells were grown in 85% Dulbecco's Modified Eagle Medium (DMEM) with 300  $\mu$ g/mL of G418 and 10% FBS under identical conditions, and harvested at 80% to 85% confluency. Total cellular RNA was purified using the Qiagen AllPrep Mini Kit according to the manufacturer's protocol. Microarray expression analysis was performed using the Human Genome U133 Plus 2.0 array (Affymetrix), and the data were normalized together using Robust Multi-array Average (RMA; Affymetrix). This generates expression values for each probe in  $\log_2$  space. We then calculated the absolute value of the relative fold-change score (Supplementary Table S8) as the following: absolute value [(7250\_PF RMA signal) - (7250\_PF Nil RMA signal)]. These values were further analyzed using the GSEA algorithm (see method below; <http://www.broad.mit.edu/gsea/>).

### Mouse Model of PAX3-FOXO1 Expression in Somite or Forelimb

Expression data for PAX3-FOXO1 expression in mouse somite and forelimb were derived from the experiments previously described (28).

### GSEA Analysis

To test whether somatically altered genes in fusion-negative rhabdomyosarcoma overlap with genes downstream of fusion-positive rhabdomyosarcoma, we performed GSEA (67) on three PAX gene fusion model systems. Gene expression from each model system was ranked according to absolute fold-change expression over the corresponding control. GSEA analysis (<http://www.broadinstitute.org/gsea/index.jsp>) was performed using default parameter settings. *P* values were calculated on the basis of Kolmogorov-Smirnov statistic with RPT.

### Disclosure of Potential Conflicts of Interest

J. Chmielecki is employed as a Senior Scientist at Foundation Medicine. M. Meyerson has ownership interest (including patents) in Foundation Medicine and is a consultant/advisory board member of the same. No potential conflicts of interest were disclosed by the other authors.

### Disclaimer

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

### Authors' Contributions

**Conception and design:** J.F. Shern, J.S. Wei, Y.K. Song, J.R. Anderson, S.X. Skapek, F.G. Barr, M. Meyerson, D.S. Hawkins, J. Khan  
**Development of methodology:** J.F. Shern, J.S. Wei, Y.K. Song, T. Badgett, G. Getz, D.S. Hawkins, J. Khan  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** J. Chmielecki, J.S. Wei, D. Auclair,



Y.K. Song, C. Tolman, H. Liao, J. Mora, J.R. Anderson, S.X. Skapek, F.G. Barr, D.S. Hawkins, J. Khan

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** J.F. Shern, L. Chen, J. Chmielecki, J.S. Wei, R. Patidar, M. Rosenberg, J. Wang, Y.K. Song, S. Zhang, D. Bogen, A.S. Brohl, S. Sindiri, F.G. Barr, M. Meyerson, D.S. Hawkins, J. Khan

**Writing, review, and/or revision of the manuscript:** J.F. Shern, L. Chen, J.S. Wei, R. Patidar, D. Bogen, A.S. Brohl, D. Catchpoole, T. Badgett, J. Mora, J.R. Anderson, S.X. Skapek, F.G. Barr, M. Meyerson, D.S. Hawkins, J. Khan

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** J.F. Shern, L. Chen, J.S. Wei, R. Patidar, L. Ambrogio, D. Auclair, L. Hurd, H. Liao, S. Sindiri, D. Catchpoole, J. Mora, J.R. Anderson, J. Khan

**Study supervision:** J.F. Shern, S.X. Skapek, J. Khan

**Project management:** L. Ambrogio

**Biobanking support:** D. Catchpoole

## Acknowledgments

The authors thank the Children's Oncology Group Soft Tissue Sarcoma Committee and the BioPathology Center, especially Dr. Julie Gastier-Foster, for their careful collection and annotation of clinical samples. The authors also thank the staff of the NCI Cancer Genomics Research Laboratory and core RNA sequencing facility for their contribution to the study. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (<http://biowulf.nih.gov>).

## Grant Support

J.F. Shern, L. Chen, J.S. Wei, R. Patidar, J. Wang, Y.K. Song, C. Tolman, L. Hurd, H. Liao, S. Zhang, D. Bogen, A.S. Brohl, S. Sindiri, T. Badgett, F.G. Barr, and J. Khan are supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. M. Meyerson is supported by M.B. Zuckerman and Team Dragonfly of the Pan-Mass Challenge. J. Chmielecki is supported by an American Cancer Society AstraZeneca Postdoctoral Fellowship. D. Catchpoole is supported by a joint appointment with the Discipline of Pediatrics and Child Health, Sydney Medical School, University of Sydney, and receives funding support through The Kids Cancer Project.

Received September 16, 2013; revised December 2, 2013; accepted December 4, 2013; published OnlineFirst January 23, 2014.

## REFERENCES

- Ognjanovic S, Linabery AM, Charbonneau B, Ross JA. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975–2005. *Cancer* 2009;115:4218–26.
- Malempati S, Hawkins DS. Rhabdomyosarcoma: review of the Children's Oncology Group (COG) soft-tissue Sarcoma committee experience and rationale for current COG studies. *Pediatr Blood Cancer* 2012;59:5–10.
- Breneman JC, Lyden E, Pappo AS, Link MP, Anderson JR, Parham DM, et al. Prognostic factors and clinical outcomes in children and adolescents with metastatic rhabdomyosarcoma—a report from the Intergroup Rhabdomyosarcoma Study IV. *J Clin Oncol* 2003;21:78–84.
- Linardic CM. PAX3-FOXO1 fusion gene in rhabdomyosarcoma. *Cancer Lett* 2008;270:10–8.
- Scrabble H, Cavenee W, Ghavimi F, Lovell M, Morgan K, Sapienza C. A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. *Proc Natl Acad Sci U S A* 1989;86:7480–4.
- Taylor AC, Shu L, Danks MK, Poquette CA, Shetty S, Thayer MJ, et al. P53 mutation and MDM2 amplification frequency in pediatric rhabdomyosarcoma tumors and cell lines. *Med Pediatr Oncol* 2000;35:96–103.
- Stratton MR, Fisher C, Gusterson BA, Cooper CS. Detection of point mutations in N-ras and K-ras genes of human embryonal rhabdomyosarcomas using oligonucleotide probes and the polymerase chain reaction. *Cancer Res* 1989;49:6324–7.
- Shukla N, Ameer N, Yilmaz I, Nafa K, Lau CY, Marchetti A, et al. Oncogene mutation profiling of pediatric solid tumors reveals significant subsets of embryonal rhabdomyosarcoma and neuroblastoma with mutated genes in growth signaling pathways. *Clin Cancer Res* 2012;18:748–57.
- Taylor JG VI, Cheuk AT, Tsang PS, Chung JY, Song YK, Desai K, et al. Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. *J Clin Invest* 2009;119:3395–407.
- Maurer HM, Beltangady M, Gehan EA, Crist W, Hammond D, Hays DM, et al. The intergroup rhabdomyosarcoma study—I. A final report. *Cancer* 1988;61:209–20.
- Barr FG, Galili N, Holick J, Biegel JA, Rovera G, Emanuel BS. Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. *Nat Genet* 1993;3:113–7.
- Davis RJ, D'Cruz CM, Lovell MA, Biegel JA, Barr FG. Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. *Cancer Res* 1994;54:2869–72.
- Wachtel M, Dettling M, Koscielniak E, Stegmaier S, Treuner J, Simon-Klingenstein K, et al. Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. *Cancer Res* 2004;64:5539–45.
- Marshall AD, Lagutina I, Grosveld GC. PAX3-FOXO1 induces cannabinoid receptor 1 to enhance cell invasion and metastasis. *Cancer Res* 2011;71:7471–80.
- Iolascon A, Faienza MF, Coppola B, Rosolen A, Basso G, Della-Ragione F, et al. Analysis of cyclin-dependent kinase inhibitor genes (CDKN2A, CDKN2B, and CDKN2C) in childhood rhabdomyosarcoma. *Gene Chromosome Cancer* 1996;15:217–22.
- Reichek JL, Duan FH, Smith LM, Gustafson DM, O'Connor RS, Zhang CN, et al. Genomic and clinical analysis of amplification of the 13q31 chromosomal region in alveolar rhabdomyosarcoma: a report from the Children's Oncology Group. *Clin Cancer Res* 2011;17:1463–73.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–81.
- Paulson V, Chandler G, Rakheja D, Galindo RL, Wilson K, Amatruda JF, et al. High-resolution array CGH identifies common mechanisms that drive embryonal rhabdomyosarcoma pathogenesis. *Gene Chromosome Cancer* 2011;50:397–408.
- Martinelli S, McDowell HP, Vigne SD, Kokai G, Uccini S, Tartaglia M, et al. RAS signaling dysregulation in human embryonal Rhabdomyosarcoma. *Genes Chromosomes Cancer* 2009;48:975–82.
- Chen YY, Takita J, Hiwatari M, Igarashi T, Hanada R, Kikuchi A, et al. Mutations of the PTPN11 and RAS genes in rhabdomyosarcoma and pediatric hematological malignancies. *Gene Chromosome Cancer* 2006;45:583–91.
- Bridge JA, Liu J, Qualman SJ, Suijkerbuijk R, Wenger G, Zhang J, et al. Genomic gains and losses are similar in genetic and histologic subsets of rhabdomyosarcoma, whereas amplification predominates in embryonal with anaplasia and alveolar subtypes. *Genes Chromosomes Cancer* 2002;33:310–21.
- Barr FG, Duan F, Smith LM, Gustafson D, Pitts M, Hammond S, et al. Genomic and clinical analyses of 2p24 and 12q13-q14 amplification in alveolar rhabdomyosarcoma: a report from the Children's Oncology Group. *Genes Chromosomes Cancer* 2009;48:661–72.
- Williamson D, Missiaglia E, de Reynies A, Pierron G, Thuille B, Palenzuela G, et al. Fusion gene-negative alveolar rhabdomyosarcoma is clinically and molecularly indistinguishable from embryonal rhabdomyosarcoma. *J Clin Oncol* 2010;28:2151–8.

## RESEARCH ARTICLE

Shern et al.

24. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011;39:D691-7.
25. Davicioni E, Finckenstein FG, Shahbazian V, Buckley JD, Triche TJ, Anderson MJ. Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. *Cancer Res* 2006;66:6936-46.
26. Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res* 1998;58:5009-13.
27. Cao L, Yu Y, Bilke S, Walker RL, Mayeenuddin LH, Azorsa DO, et al. Genome-wide identification of PAX3-FKHR binding sites in rhabdomyosarcoma reveals candidate target genes important for development and cancer. *Cancer Res* 2010;70:6497-508.
28. Lagha M, Sato T, Regnault B, Cumano A, Zuniga A, Licht J, et al. Transcriptome analyses based on genetic screens for Pax3 myogenic targets in the mouse embryo. *BMC Genomics* 2010;11:696.
29. Skapek SX, Anderson J, Barr FG, Bridge JA, Gastier-Foster JM, Parham DM, et al. PAX-FOXO1 fusion status drives unfavorable outcome for children with rhabdomyosarcoma: a children's oncology group report. *Pediatr Blood Cancer* 2013;60:1411-7.
30. Missiaglia E, Williamson D, Chisholm J, Wirapati P, Pierron G, Petel F, et al. PAX3/FOXO1 fusion gene status is the key prognostic molecular marker in rhabdomyosarcoma and significantly improves current risk stratification. *J Clin Oncol* 2012;30:1670-7.
31. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. *Nat Genet* 2013;45:279-84.
32. Grossmann V, Tiacchi E, Holmes AB, Kohlmann A, Martelli MP, Kern W, et al. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. *Blood* 2011;118:6153-63.
33. Zhang JH, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* 2012;481:329-34.
34. Pugh TJ, Weeraratne SD, Archer TC, Krummel DAP, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 2012;488:106-10.
35. Lagutina I, Conway SJ, Sublett J, Grosveld GC. Pax3-FKHR knock-in mice show developmental aberrations but do not develop tumors. *Mol Cell Biol* 2002;22:7204-16.
36. Linardic CM, Naini S, Herndon JE, Kesserwan C, Qualman SJ, Counter CM. The PAX3-FKHR fusion gene of rhabdomyosarcoma cooperates with loss of p16(INK4A) to promote bypass of cellular senescence. *Cancer Res* 2007;67:6691-9.
37. Keller C, Arenkiel BR, Coffin CM, El-Bardeesy N, DePinho RA, Capocchi MR. Alveolar rhabdomyosarcomas in conditional Pax3-FKHR mice: cooperativity of Ink4a/ARF and Trp53 loss of function. *Genes Dev* 2004;18:2614-26.
38. Ascierto PA, Schadendorf D, Berking C, Agarwala SS, van Herpen CML, Queirolo P, et al. MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study. *Lancet Oncol* 2013;14:249-56.
39. Renshaw J, Taylor KR, Bishop R, Valenti M, De Haven Brandon A, Gowan S, et al. Dual blockade of the PI3K/AKT/mTOR (AZD8055) and RAS/MEK/ERK (AZD6244) pathways synergistically inhibits rhabdomyosarcoma cell growth *in vitro* and *in vivo*. *Clin Cancer Res* 2013;19:5940-51.
40. Huynh KD, Fischle W, Verdin E, Bardwell VJ. BCoR, a novel corepressor involved in BCL-6 repression. *Genes Dev* 2000;14:1810-23.
41. Gearhart MD, Corcoran CM, Wamstad JA, Bardwell VJ. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol* 2006;26:6880-9.
42. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao YJ, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 2012;486:395-9.
43. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011;476:298-303.
44. Petricoin EF III, Espina V, Araujo RP, Midura B, Yeung C, Wan X, et al. Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. *Cancer Res* 2007;67:3431-40.
45. Cen L, Arnoczky KJ, Hsieh FC, Lin HJ, Qualman SJ, Yu SL, et al. Phosphorylation profiles of protein kinases in alveolar and embryonal rhabdomyosarcoma. *Mod Pathol* 2007;20:936-46.
46. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;327:78-81.
47. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* 2012;481:157-63.
48. Reumers J, De Rijk P, Zhao H, Liekens A, Smeets D, Cleary J, et al. Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat Biotechnol* 2011;30:61-8.
49. Molenaar JJ, Koster J, Ebus ME, Zwiijnenburg DA, van Sluis P, Lamers F, et al. The genomic landscape of neuroblastoma. *Cell Oncol* 2012;35:S11-S.
50. Carnevali P, Baccash J, Halpern AL, Nazarenko I, Nilsen GB, Pant KP, et al. Computational techniques for human genome resequencing using mated gapped reads. *J Comput Biol* 2012;19:279-92.
51. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:1-9.
52. Sorensen PHB, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. *J Clin Oncol* 2002;20:2672-9.
53. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105-11.
54. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* 2011;12:1-15.
55. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* 2011;7:1-16.
56. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010;28:511-5.
57. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE* 2009;4:e7767.
58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-303.
59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491-8.
60. Barbieri CE, Baca SC, Lawrence MS, Demicheli F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 2012;44:685-9.
61. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* 2011;333:1157-60.
62. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011;471:467-72.
63. Wei XM, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, et al. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* 2011;43:442-6.
64. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B Met* 1995;57:289-300.

## The Genomic Landscape of Rhabdomyosarcoma

## RESEARCH ARTICLE

65. Diskin SJ, Eck T, Greshock J, Mosse YP, Naylor T, Stoeckert CJ, et al. STAC: a method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res* 2006;16:1149–58.
66. Khan J, Bittner ML, Chen YD, Faller AJ, Saal LH, Azorsa DA, et al. Elucidation of the downstream targets of the PAX3-FKHR fusion oncogene found in alveolar rhabdomyosarcoma using cDNA microarrays. *Pediatr Res* 1999;45:148A.
67. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.